Title: Robustness checks appendix for "Hide and seek: The connection between false beliefs and perceptions of government transparency"
Authors: Mathieu Lavigne (1), Éric Bélanger (1), Richard Nadeau (2), Jean-François Daoust (3), Erick Lachapelle (1)
Date: March 16th, 2022
Note: The material contained herein is supplementary to the article named in the title and published in the Harvard Kennedy School (HKS) Misinformation Review.

Appendix C: Robustness checks

This Appendix presents the results of the robustness checks that were made to evaluate the robustness of the results. First, we show that the results are similar when using entropy balancing to make individuals who assigned high and low levels of truthfulness to COVID-19 false statements more comparable on relevant observable covariates (i.e., variables that are associated both with the perceived truthfulness of false statements and with perceptions of transparency). The results are presented in Figure C1 and Table C1. Second, we examine how each false belief item contributes to perceptions of transparency and show that false beliefs that are conspiratorial in nature and false beliefs that are not necessarily related to conspiracy theories both predict perceptions of transparency, as reported in Tables C2 and C3. Relatedly, we used exploratory factor analysis to examine the structure of false beliefs and assess if there could exist different types of false beliefs. We identify two types of false beliefs among the false statements we used—those related to the severity of COVID-19 and those related to the origin of the virus—and show their respective association with perceptions of transparency. The results of these analyses are presented in Tables C4 to C7. We also examined whether the results remain unchanged when including both true and false statements in the false beliefs index, as reported in Table C8. Figure C2 shows the results of a specification curve analysis that we conducted to examine how much the size of the coefficient for "false beliefs" depends on the specification of our OLS model. A simulation was conducted where the model could include any combination of variables among a list of 22 control variables. Finally, Figure C3 shows that the results of our multinomial logistic regression examining the relationship between false beliefs and categories of answers to an open-ended question asking respondents to explain what governments are hiding about the pandemic are not influenced by the method used to deal with disagreements between coders.

Entropy balancing

Entropy balancing, a preprocessing method that reweights observations to achieve covariate balance (i.e., to remove observable differences between treated and control units), was used to better isolate the effect of false beliefs from that of other correlated variables, by making sure that we compare individuals that are similar on other relevant covariates (Hainmueller, 2012). Given the continuous nature of the false-beliefs variable, we used entropy balancing for continuous treatment, which reweights units to achieve zero correlations between the treatment variable and covariates (Tübbicke, 2020). Using standard entropy balancing (which requires dichotomizing the treatment based on whether respondents have endorsed at least one of the false statements or not) or also including trust in the federal and provincial governments in the entropy balancing yield similar results. Entropy balancing succeeded in removing any correlation between false beliefs and the covariates, as reported in Figure C1. Entropy balancing also successfully removed the correlation between false beliefs and the square of each variable. Using entropy balancing does not change our conclusions about the association between false beliefs and perceptions of transparency, as coefficients are of a similar magnitude, as shown in Table C1.

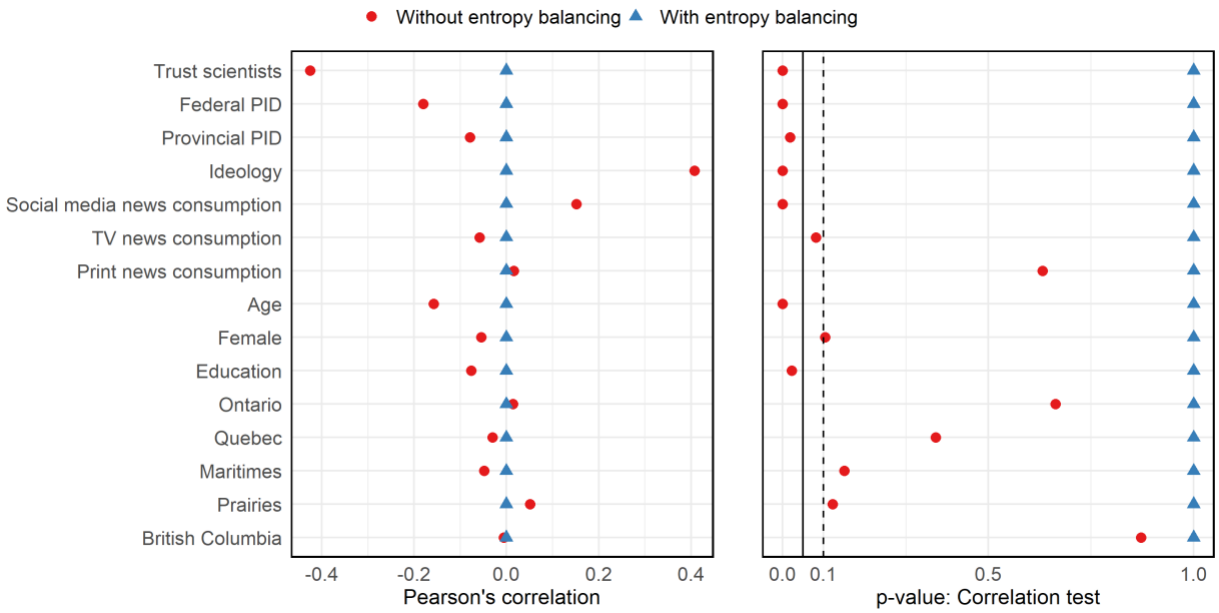

Figure C1. Pearson's correlation between false beliefs and covariates before and after entropy balancing. *Left panel shows Pearson's correlation coefficients between false beliefs and covariates. A correlation coefficient of zero indicates there is no linear relationship between false beliefs and covariates, correlations greater than zero indicates that false beliefs are associated with higher values on the covariates, and correlations lower than zero indicates that false beliefs are associated with lower values on the covariates. The right panel shows the p-value for a covariate-by-covariate correlation test. A p-value lower than 0.05 indicates that the correlation coefficient is significant.*

Note that we balanced partisan identification, ideology, and trust in scientists at t-1 in the model predicting *changes* in perceptions that governments are hiding information. The correlations are also reduced to 0. Measuring these variables at t instead yield similar results.

Table C1. *Results of OLS regressions examining the relationship between false beliefs and perceptions of transparency after entropy balancing. Unstandardized regression coefficients are shown with robust (HC2) standard errors in parentheses.*

	Hides information	Lacks transparency	Δ Hides information
	(1)	(2)	(3)
False beliefs	0.66***	0.30***	3.48***
	(0.05)	(0.05)	(0.58)
Trust scientists	-0.21**	-0.11+	
	(0.07)	(0.06)	
Federal PID	-0.01	-0.004	
	(0.03)	(0.03)	
Provincial PID	0.03	-0.08**	
	(0.03)	(0.03)	
Left-right ideology	-0.01	-0.11	
	(0.07)	(0.07)	
Trust scientists (t-1)			-1.65*

			(0.82)
Federal PID (t-1)			-0.08
			(0.29)
Provincial PID (t-1)			-0.16
			(0.29)
Left-right ideology (t-1)			0.22
			(0.53)
Social media news consumption	-0.03	-0.03	-0.05
	(0.04)	(0.04)	(0.37)
TV news consumption	0.02	-0.03	-0.21
	(0.04)	(0.05)	(0.41)
Print news consumption	-0.002	0.03	0.62
	(0.05)	(0.06)	(0.44)
Age	-0.10^{+}	-0.07	-0.93^{+}
	(0.06)	(0.06)	(0.52)
Female	0.004	-0.04^{+}	-0.33
	(0.02)	(0.03)	(0.24)
Education	0.02	-0.03	-0.32
	(0.03)	(0.03)	(0.35)
Hides information (t-1)			-0.46***
			(0.05)
Regional fixed effects	Yes	Yes	Yes
Observations	880	848	535
R^2	0.35	0.16	0.24
Adjusted R^2	0.34	0.14	0.22
Residual Std. Error	0.25	0.29	2.11

+ p < .1; * p < .05; ** p < .01; *** p < .001

Disaggregating the false beliefs index into its constitutive elements

We did different tests to examine how specific items contribute to the relationship between false beliefs and perceptions of government transparency. First, it is interesting, from a theoretical perspective, to distinguish between items that are expressly conspiratorial (e.g., they identify powerful groups that can benefit from some secret plans) and those that are not expressly conspiratorial (e.g., health-related misinformation, which might still have been used as part of some conspiracy theories). The three conspiratorial items are as follows: (1) the government is exaggerating the risks of the coronavirus to be able to restrict people's rights and freedoms; (2) the virus was created by China to increase its power in the world; and (3) the virus has been created by large corporations because some of them can directly profit from it. The three health-related statements are as follows: (1) the prolonged use of masks can lead to CO_2 intoxication or oxygen deficiency; (2) hydroxychloroquine is an effective treatment against COVID-19; and (3) coronavirus figures are inflated because a significant number of people tested positive are not infected with the virus.

The first two columns of Table C2 and C3 respectively present the disaggregated results of the conspiratorial and health-related (not-expressly-conspiratorial) items, while the third column presents the results when including all items. The results suggest that the belief that the government is exaggerating the pandemic to be able to restrict people's rights, that figures are inflated because of false positives, and that China has created the coronavirus to increase its power in the world have the strongest independent effects on perceptions of transparency.

Table C2. *Results of OLS regressions examining the relationship between the perceived truthfulness of each false statement and perceptions that governments are hiding information about the pandemic. Unstandardized regression coefficients are shown with robust (HC2) standard errors in parentheses.*

	Hides information	Hides information	Hides information
	(1)	(2)	(3)
Restrict rights	0.39***		0.28***
	(0.05)		(0.06)
Power China	0.18***		0.16**
	(0.05)		(0.05)
Benefit companies	0.02		0.01
	(0.06)		(0.06)
CO_2 masks		0.11*	-0.0001
		(0.04)	(0.05)
Hydroxychloroquine		0.03	-0.04
		(0.04)	(0.04)
False positives		0.39***	0.20***
		(0.05)	(0.06)
Trust scientists	-0.04*	-0.05*	-0.04+
	(0.02)	(0.02)	(0.02)
Federal PID	0.003	0.005	0.004

	(0.02)	(0.02)	(0.02)
Provincial PID	-0.02	0.02	-0.03
	(0.05)	(0.06)	(0.05)
Left-right ideology	0.01	-0.01	-0.003
	(0.03)	(0.03)	(0.03)
Social media news consumption	-0.03	-0.02	-0.02
	(0.03)	(0.03)	(0.03)
TV news consumption	-0.02	-0.03	-0.02
	(0.04)	(0.04)	(0.04)
Print news consumption	-0.08*	-0.09*	-0.09*
	(0.04)	(0.04)	(0.04)
Age	0.02	0.01	0.02
	(0.02)	(0.02)	(0.02)
Female	-0.03	-0.01	-0.02
	(0.02)	(0.03)	(0.03)
Education	-0.06	-0.10*	-0.05
	(0.05)	(0.05)	(0.05)
Regional fixed effects	Yes	Yes	Yes
Observations	880	880	880
R^2	0.40	0.36	0.41
Adjusted R^2	0.39	0.35	0.40
Residual Std. Error	0.23	0.24	0.23

+ p < .1; * p < .05; ** p < .01; *** p < .001

Table C3. *Results of OLS regressions examining the relationship between the perceived truthfulness of each false statement and changes in perceptions that governments are hiding information about the pandemic between Wave 2 (June 2020) and Wave 3 (January 2021). Unstandardized regression coefficients are shown with robust (HC2) standard errors in parentheses.*

	Δ Hides information	Δ Hides information	Δ Hides information
	(1)	(2)	(3)
Restrict rights	2.77***		2.34***
	(0.56)		(0.66)
Power China	0.33		0.34
	(0.52)		(0.53)
Benefit companies	0.76		0.71
	(0.52)		(0.57)
CO_2 masks		0.82	-0.03
		(0.52)	(0.55)
Hydroxychloroquine		0.09	-0.44
		(0.46)	(0.47)
False positives		2.35***	0.94
		(0.57)	(0.66)
Trust scientists (t-1)	-1.28*	-1.23*	-1.23*
	(0.56)	(0.57)	(0.57)
Federal PID (t-1)	-0.12	-0.17	-0.11
	(0.28)	(0.28)	(0.29)
Provincial PID (t-1)	-0.02	-0.09	-0.05
	(0.27)	(0.27)	(0.27)
Left-right ideology (t-1)	0.19	0.44	0.22
	(0.48)	(0.49)	(0.49)
Social media news consumption	-0.06	-0.10	-0.07
	(0.33)	(0.33)	(0.33)
TV news consumption	-0.12	-0.12	-0.11
	(0.36)	(0.35)	(0.36)
Print news consumption	0.29	0.35	0.32
	(0.37)	(0.38)	(0.38)
Age	-0.29	-0.38	-0.37
	(0.46)	(0.47)	(0.46)
Female	0.01	-0.01	0.02
	(0.20)	(0.21)	(0.21)

Education	-0.03	0.06	-0.03
	(0.27)	(0.29)	(0.28)
Hides information (t-1)	-0.54***	-0.49***	-0.55***
	(0.05)	(0.05)	(0.05)
Regional fixed effects	Yes	Yes	Yes
Observations	535	535	535
R^2	0.30	0.27	0.31
Adjusted R^2	0.28	0.24	0.28
Residual Std. Error	1.95	2.00	1.95

+ p < .1; * p < .05; ** p < .01; *** p < .001

Exploratory factor analysis

We also ran an exploratory factor analysis[1] to see if different COVID-19 false statements can potentially load on different factors and have a different impact on perceptions of government transparency. We still believe that it is justified to use a single factor given the high correlations between items, the high Cronbach's alpha, and the fact that principal component analysis indicates that there is a single factor with an Eigenvalue greater than 1 (4.2). Nevertheless, the results of the exploratory factor analysis presented in Table C4 suggest that some items might contribute more to some subdimensions than others. The results suggest that two factors are sufficient, since we fail to reject the null hypothesis that two factors are sufficient, X^2 (4, N = 1005) = 1.09, p = .90.

[1] Exploratory factor analysis is a statistical method that is used to identify the latent constructs (i.e., factors) that underly a set of variables (Watkins, 2018). We use exploratory factor analysis to reduce our false beliefs items to a smaller number of hypothetical constructs that are assumed to explain the order and structure among false beliefs.

***Table C4**. Factor loadings and communality for varimax rotated two-factor solution for six false beliefs items.*

	Factor loading		
	1 Severity	2 Origin	Communality
The government is exaggerating the risks of the coronavirus to be able to restrict people's rights and freedoms.	0.76	0.45	0.78
Coronavirus figures are inflated because a significant number of people tested positive are not infected with the virus.	0.79	0.39	0.78
The virus was created by China to increase its power in the world.	0.37	0.73	0.67
The virus has been created by large corporations because some of them can directly profit from it.	0.44	0.76	0.77
The prolonged use of masks can lead to CO_2 intoxication or oxygen deficiency.	0.61	0.53	0.65
Hydroxychloroquine is an effective treatment against COVID-19.	0.50	0.48	0.49

***Table C5**. Eigenvalues, percentages of variance, and cumulative percentages for factors for six false beliefs items.*

Factor	Eigenvalue	% of variance	Cumulative %
1	4.20	36.57	36.57
2	0.51	32.49	69.05

Based on the results of the factor analysis, we created two factors using the factor loadings, that is weighting each item based on their contribution to each factor. We have named the first factor the "Severity factor," given that the two items that contribute the most to this factor are the idea that (1) the government is exaggerating the risks of the coronavirus to be able to restrict people's rights and freedoms; and (2) coronavirus figures are inflated because a significant number of people tested positive are not infected with the virus. We named the second factor the "Origin factor," given the comparatively high contribution of the following two items: (1) the virus was created by China to increase its power in the world; (2) the virus has been created by large corporations because some of them can directly profit from it. Tables C6 and C7 show the results of the regression analysis when using these two factors in place of the false-beliefs index. The results suggest that different types of false beliefs (about the severity and origin of the pandemic) can have different, independent effects on perceptions of transparency.

Table C6. *Results of OLS regressions examining the relationship between different false beliefs indices and perceptions that governments are hiding information about the pandemic. Unstandardized regression coefficients are shown with robust (HC2) standard errors in parentheses.*

	Hides information	Hides information	Hides information
	(1)	(2)	(3)
Severity factor	0.16***		0.15***
	(0.02)		(0.01)
Origin factor		0.12***	0.10***
		(0.01)	(0.01)
Trust scientists	-0.14**	-0.17**	-0.06
	(0.05)	(0.05)	(0.05)
Federal PID	-0.05*	-0.06*	-0.05*
	(0.02)	(0.02)	(0.02)
Provincial PID	0.004	-0.01	0.01
	(0.02)	(0.02)	(0.02)
Left-right ideology	0.06	0.13*	-0.02
	(0.06)	(0.06)	(0.05)
Social media news consumption	0.01	0.03	-0.01
	(0.03)	(0.03)	(0.03)
TV news consumption	-0.002	-0.07*	-0.02
	(0.04)	(0.04)	(0.03)
Print news consumption	-0.02	-0.02	-0.03
	(0.04)	(0.04)	(0.04)
Age	-0.10*	-0.09*	-0.07+
	(0.04)	(0.04)	(0.04)
Female	0.01	0.004	0.02
	(0.02)	(0.02)	(0.02)
Education	-0.03	-0.01	-0.02
	(0.03)	(0.03)	(0.03)
Regional fixed effects	Yes	Yes	Yes
Observations	879	879	879
R^2	0.34	0.26	0.40
Adjusted R^2	0.33	0.24	0.38
Residual Std. Error	0.24	0.26	0.23

+ p < .1; * p < .05; ** p < .01; *** p < .001

Table C7. *Results of OLS regressions examining the relationship between two false beliefs indices and changes in perceptions that governments are hiding information about the pandemic between Wave 2 (June 2020) and Wave 3 (January 2021). Unstandardized regression coefficients are shown with robust (HC2) standard errors in parentheses.*

	ΔHides information	ΔHides information	ΔHides information
	(1)	(2)	(3)
Severity factor	1.00***		0.99***
	(0.19)		(0.17)
Origin factor		0.60***	0.59***
		(0.14)	(0.16)
Trust scientists (t-1)	-1.41*	-1.78**	-1.12*
	(0.55)	(0.56)	(0.57)
Federal PID (t-1)	-0.16	-0.20	-0.15
	(0.28)	(0.30)	(0.28)
Provincial PID (t-1)	-0.08	-0.10	-0.04
	(0.27)	(0.29)	(0.27)
Left-right ideology (t-1)	0.43	0.81	0.21
	(0.50)	(0.52)	(0.48)
Social media news consumption	-0.07	0.25	-0.15
	(0.35)	(0.34)	(0.33)
TV news consumption	-0.03	-0.35	-0.11
	(0.37)	(0.37)	(0.36)
Print news consumption	0.39	0.18	0.32
	(0.38)	(0.39)	(0.38)
Age	-0.41	-0.14	-0.30
	(0.47)	(0.50)	(0.47)
Female	-0.003	-0.08	0.005
	(0.21)	(0.21)	(0.20)
Education	-0.02	0.22	0.01
	(0.29)	(0.28)	(0.28)
Hides information (t-1)	-0.47***	-0.43***	-0.53***
	(0.06)	(0.04)	(0.05)
Regional fixed effects	Yes	Yes	Yes
Observations	535	535	535
R^2	0.26	0.20	0.29
Adjusted R^2	0.24	0.18	0.27
Residual Std. Error	2.00	2.07	1.96

+ p < .1; * p < .05; ** p < .01; *** p < .001

Accounting for the acceptance or rejection of true statements

In the survey, we also measured the perceived truthfulness of three statements that are assumed to be true, based on the best expert evidence available at the time. These statements were included to avoid bias caused by asking respondents to only evaluate the truthfulness of false statements. The three statements are as follows:

- The virus is the result of an accidental animal-human transmission that occurred in China (Maxmen & Mallapaty, 2021).
- The only permanent solution to this pandemic is developing a vaccine (van Riel & de Wit, 2020).
- The number of accidental poisonings involving hand sanitizer and children has increased significantly since the pandemic began (Feireisen, 2020).

To incorporate the effect of accepting or rejecting true information into our assessment of the relationship between false beliefs and perceptions of transparency, we computed a new variable consisting of the difference between the perceived truthfulness of the false and true statements. The variable is measured on a -1 to 1 scale, where -1 indicates a complete belief in true statements and disbelief in false statements and 1 indicates the reverse. Including the acceptance of true statements in our measure of false beliefs does not change our conclusions. The results presented in Table C8 show that the greater the distance between the perceived truthfulness of false and true statements, the greater the likelihood of believing that governments are hiding information or lacking transparency about what motivates their decisions. To put it simply, those who believe the COVID-19 misinformation statements but not the factual statements are more likely to have negative perceptions of government transparency.

Table C8. *Results of OLS regressions examining the relationship between the difference in the perceived truthfulness of the false and true statements (-1 to 1) and perceptions of transparency. Unstandardized regression coefficients are shown with robust (HC2) standard errors in parentheses.*

	Hides information	Lacks transparency	ΔHides information
	(1)	(2)	(3)
False - true beliefs	0.42***	0.24***	3.00***
	(0.04)	(0.04)	(0.40)
Trust scientists	-0.12*	0.02	
	(0.05)	(0.07)	
Trust scientists (t-1)			-1.35**
			(0.56)
Federal PID	-0.04	-0.01	-0.11
	(0.02)	(0.03)	(0.29)
Provincial PID	-0.01	-0.11***	-0.10
	(0.02)	(0.03)	(0.29)
Left-right ideology	0.08	0.04	0.52
	(0.06)	(0.06)	(0.50)
Social media news consumption	0.03	-0.01	0.04
	(0.03)	(0.04)	(0.32)
TV news consumption	-0.02	-0.06	-0.07

	(0.04)	(0.04)	(0.35)
Print news consumption	0.01	0.02	0.44
	(0.04)	(0.05)	(0.37)
Age	-0.08^{+}	-0.04	-0.37
	(0.05)	(0.05)	(0.49)
Female	0.01	-0.04	-0.17
	(0.02)	(0.02)	(0.20)
Education	-0.02	-0.03	-0.03
	(0.03)	(0.03)	(0.29)
Hides information (t-1)			-0.49***
			(0.05)
Regional fixed effects	Yes	Yes	Yes
Observations	880	848	535
R^2	0.30	0.16	0.26
Adjusted R^2	0.29	0.15	0.24
Residual Std. Error	0.25	0.29	2.00

+ p < .1; * p < .05; ** p < .01; *** p < .001

Specification curve analysis

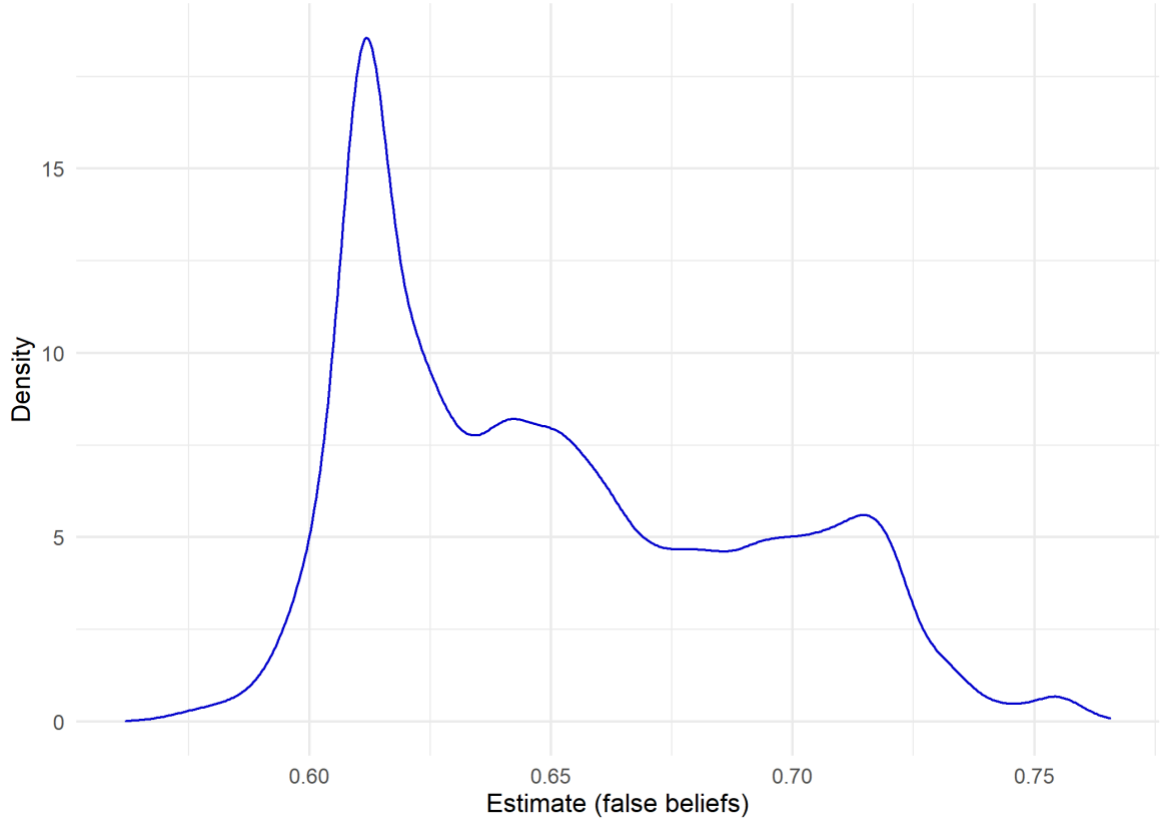

Figure C2. Density of the "False beliefs" unstandardized regression coefficient across 50,000 randomly sampled model specifications. *Results based on OLS regression models where perceptions that governments are hiding information about the pandemic is the dependent variable. Across specifications, control variables could randomly include any combination of the following variables: age, sex, education, identification with the governing federal and provincial parties, ideology, trust in scientists, frequency of exposure to COVID-19 information on television, radio, newspapers, and social media, trust in the federal and provincial governments, identification with opposition parties, generalized social trust, perceived threat of becoming unemployed, life satisfaction, and emotions caused by COVID-19 (fear, anger).*

Using different methods to deal with disagreement between coders

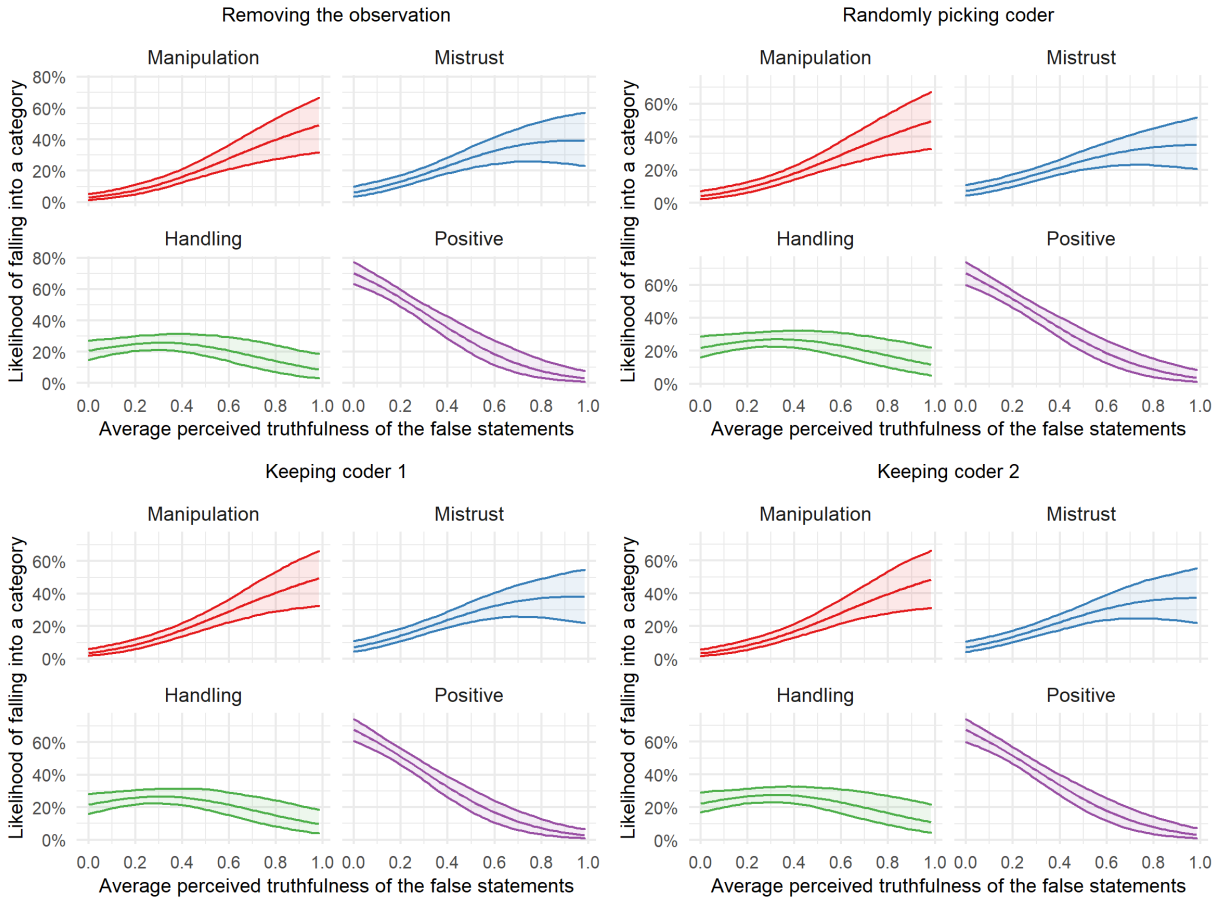

Figure C3: Results of the multinomial logistic models based on the method used to deal with disagreement between coders. *Predicted probability of falling into each response category to an open-ended question asking respondents to explain what governments are hiding about the pandemic, based on the perceived truthfulness of COVID-19 false statements. Predicted probabilities are generated from a multinomial logistic regression, with 95% confidence intervals, using an observed value approach (MNL_pred package in R). The model controls for trust in scientists, identification with the federal and provincial governing party, ideology, news consumption, and socio-demographic variables (age, gender, education, region). Four methods of dealing with disagreement are compared: removing the observation (top-left panel); randomly selecting one of the two codes (top-right panel); keeping the code of the first coder (bottom-left panel); and keeping the code of the second coder (bottom-right panel).*